

Supporting Information

© Wiley-VCH 2011

69451 Weinheim, Germany

## **The Tn Antigen—Structural Simplicity and Biological Complexity**

*Tongzhong Ju, Vivianne I. Otto, and Richard D. Cummings\**

anie\_201002313\_sm\_miscellaneous\_information.pdf

C1GALT1-cDNA 1 ATGGCCTCTAAATCCTGGCTGAATTTTAACTTCCTCTG--TGGATCAGCAATAGGATTTCTT-----TATATGTTTC 71  
pC1GALT-1 1 ATGGCCTCTAAATCCTGGCTGAATTTTAACTTCCTCTA--TGGATCGGCAATAGGGTTTATT-----TTATTTTC 71  
pC1GALT-2 1 ATGGCCTCTAAATCCTGGCTGAATTTTAACTTCCTCTG--TGGATCGGCAATAGGATTTTGTGTGTTTTTTTTTTT--TAAACGTTTC 87  
pC1GALT-3 1 TTTGTATTTTATGTTGGAGACGCGGGTTTACCATGTGTTGGCAGGATATATCTCGGTCTCCTGACCTCGTGAT-----CAGCTGCTC 78  
pC1GALT-4 1 ----ATGCACTTTCAGGAAAACAACAAAAAAGCTGGCTAC-----ATCTTTT-----TACCTTTT 51

I

C1GALT1-cDNA 72 TCAGCTATTTAGTATTTTGTGTTGGGAGAAAGGTGACACCCAGCCCTAATGTTCTTCATAATGATCCTCATGCAGGCCATTCAGATGATAA 161  
pC1GALT-1 72 TCAGCTACTTAGTATTTTGTGTTGGGAGAAAGGGTGACACCCAGCCTAATGTTCTTCATAATGATCCTCATGCGAGGCATTCAGATGATAA 161  
pC1GALT-2 88 TCAGCTATTTAGTATTTTGT--GGAGAAAGGTGACACCCAGCCCTAATGTTCTTCATAATGATCCTCATGCGAGGCATTCAGATGATAA 174  
pC1GALT-3 79 TCGGCTCTCCCAAGTGGCTGGGATTAAGGTGTGAAGCACCACCTCCTGTC--CCATTATAAATTAATCTCTTACGTAGAGGCCTTCAGATGATAA 167  
pC1GALT-4 52 TCTCTCTAATTAATTCAGTATTGGAAGAACAGTCTGACACCTAGCCCTGCTAAATCTTGTGTAACGATATCTCATGTTAAGACATTCAGTGTATAA 14

↓

C1GALT1-cDNA 162 TGGACAGAAATCATCTAGAAGGACAAATGAACTTCAATGCAGATTCTAGCCAACTAAAGATGAGAACACAGACATGCTGAAAAACCTCTA 251  
pC1GALT-1 162 TGGACAGAAATCATCTAGGAGGACAAATGAACTTCAATGCAGATTCTAGCCAACTGTAAGATGAGAACACAGACATCGCTGAAAAACCTCTA 251  
pC1GALT-2 175 TGG-CAAGATCATCTAGAAGGACAAATGAACTTCAATGCTGATTCTAGCCAACTAAAGATGAGAACACAGACATCGCTGAAAAACCTCTA 263  
pC1GALT-3 168 TGGGACAGAAATCATCTAATAAGACCTGATGAAAGTTGATGCAATATGCTAGTCACTCATAAATACCCAGAACACAGGTGTCTACTGACAACTCTA 257  
pC1GALT-4 142 TGGGCAATAACATGTAGAGAGA-TAATGAACTTCAACGGAGATGCTACCCAAATCAAGATGGGAACACAGATGTATTGAAAAAGCT--G 228

C1GALT1-cDNA 252 TCAGAAAGTTAGCAATTCTTTGCTGGGTTATGACCGGCCCTCAAAACCTAGAGAAAAAGGC-CAAAACGTCAAAGCTACTTGGGGCCAGC 340  
pC1GALT-1 252 TTAGCAAGTTAAAAATCTTTGCTGGGTTATGACAGGCCTCTCAAAACCTAGAGAAAAAGGC-CAAAACGTCAAAGCTACTTGGGGCCAGC 340  
pC1GALT-2 264 TCAGAAAGTTAAAAATCTTTGCTGGGTTATGACAGGCCCTTAAAAACCTAGAGAAAAAGGC-CAAAACCTATCAAGTTTACATGGGGCCAGC 352  
pC1GALT-3 258 TCAAAAGATGAAAAATCTTTGCTGGGTTATGACAGGCCTCAAAACCTAGAAAAAAAGATCAAGACGATCAAGAGATACATGGGGCCAGC 347  
pC1GALT-4 229 TCAGAAATGTTAAATCTTCTTTGGGTTATGACAGGCCTCAACATTTAGAGAAAAAGGC-CAAACATATCATAGCTACATGGGGCCAGC 317

C1GALT1-cDNA 341 GTTGTAAACAAAGTGTGTTTATGAGTTTCAAGAAAAATAAAGACTTCCCTGCTGTGGGATGAAAAACAAAGAGGCAGAGATCAACTAT 430  
pC1GALT-1 341 GTTGTCTAAAGTATTTTATGAGTTTCAAGAAAAATAAAGACTTCCCTGCTGTGTGGATGAAAAACAAAGAGGCAGAGATGAGCTAT 430  
pC1GALT-2 353 ATTGTCTAAAGTGT--TTATGAGTTTCAAGAAAAATAAAGACTTCCCTACTGTGGGATGAAAAACAAAGAGGCAGAGATCAGCTAT 438  
pC1GALT-3 348 GTTGCAATAAAGCGTTGTTTATGAGCTCAAGAAAAATAAAGACTTCTACTGTGGGATTAACACCAAGAAAGACAGAAACCACTGTT 437  
pC1GALT-4 318 AGTATACTAAATGT-----TGTAACACTTCTGTAGTCTGTAGTTCAGAAACCACTTAAG-ACAGCCAGCTAT 385

C1GALT1-cDNA 431 ACTGAAAAACAATTAAGCTTTTTCAGTATGTTTCATGAACATATTTAGAAGATGCTGATTGGTTTTTGAAGCAGATGATGACACGCT--- 517  
pC1GALT-1 431 ACTGAAAAACAATTAAGCCTTTTTCAGTATGTTTCATGAACATGATTTAGAAGATGCTGATTGGTTTTTGAAGCAGATGATGACACAT--- 517  
pC1GALT-2 439 ACTGAAAAACA---AGCTTTTTCAGTATGTTTCATGAACATATTTAGAAGATGCTGATTGGCTTTTGAAGCAGATGATGACACGCT--- 521  
pC1GALT-3 438 CTTGAAAAATAGTTAAAGCTTTTTCAGTATGTTTCATGACATATGTTGGAATCATGGAAGCAGATGATGATGATATATGTA 527  
pC1GALT-4 386 CTTGAAAAATAGCTA-TAGTCTTCAAAATGTTGATGACAGTTATTTAGAAGATGCTGATTGGTTCAAGGAACAGATTAACAGTGCAT--- 471

II

C1GALT1-cDNA 518 --ATGTCATACTAGACAATTTGAGCTGGCTTCTTTCAAAATAAGACCTGAAGAACCATTACTTTGGGAGAAAGATTTAAGCCCTTAT- 603  
pC1GALT-1 518 --ATGTCATACTAGATTAATTTGAGATGGCTTCTTTCAAAATAAGATGTTCTTAAGAAACATATTTATTTAGGAGAAAGATTTAAGCCCTT- 608  
pC1GALT-2 522 --ATGTCATACTAGACAATTTGAGATGGCTTCTTTCAAAATAATACCTGAAGAACCATTACTTTGGGAGAAACATTAAGCCCTATA- 608  
pC1GALT-3 528 TATAATATCAATTTGGAACAATTGAAATGAGCTTCTCAACAATATTAACCTGTAAGAACTTCACTTTGGGAAAGATTTAAGCACTGTC- 616  
pC1GALT-4 472 --ATGTCGTACTAGACAACCTTGAGATCACTGCTTTCAAACTATGAGCCTTAAGAACCATTCTCTTACAGAGATG--TAAGCCCTGT- 554

C1GALT1-cDNA 604 -GTAAAGCAGGGCTACAT-GAGTGGAGGAGCAGGATATGTACTAAGCAAAGAA-GCCTTGAAAGAGATTTGTTGATGCATTTAAAAACAGAC 690  
pC1GALT-1 604 -GTAAAGCAGGGCTGTAT-GAGTGGAGGAGCAGGATATTTACTAAGCAAAGAAAGCCTTGAAAGAGATTTCTTGTATGCATTTAAAAACAGAC 691  
pC1GALT-2 609 GTTAAAGCAGGGCTACATTTGAGTGGAGGAGCAGGATATGATATGTTCTAAGCAAAGAA-GCCTTGAAAGAGATTTGTTGATGCATTTGAAACAGAC 697  
pC1GALT-3 617 -AGAAACAGGACTACAT-GAGTGGAGGAGCAGGATATGTACTGAGCAAAGAA-GCCTCAAGGAGATTAATTTGTTGTTGGTCAAAACAA 703  
pC1GALT-4 555 -GGAAATTAAGCTGTAT-AGTGGAGAGCAGGATATGAGCTGAACAAATTA-TCCTCAGAGTGATTTGGTTACTGTTATTAACCT--- 638

C1GALT1-cDNA 691 AAGTGTAACAATAGTTCTCTCCAATGAAGACTTAGCACTGGGGAGATGCATGGAATTTATGAATGTAAGAGCAGGAGATTCAGAGATACCC 780  
pC1GALT-1 692 AAGTGT-CAATAGTTCTCTCCAATGAAGACTTAGCACTGGGGAGATGCATGGAATTTATAAATGTAAGAGCAGGAGATTCAGAGATCCC 780  
pC1GALT-2 698 AAGTGTAACAATAGTTCTCTCCAATGAAGACT--AGCACTGGGGAGATGCATGGAATTTATAAATGTAAGAGCAGGAGATTCAGAGATACC 785  
pC1GALT-3 704 ACGGGCACAAGCTGTGTCTTCTCTGAATTAATGGCACTGGGGAGATGCATGGAATTTGTAAGGTTGGAACAGAGAGGATTCAGATATCAC 793  
pC1GALT-4 639 -----ATTATTCTCTTACAGAGGATCAGAGTTGCTGTTCTGATCTTGCAGTTTGTATTCTTGACCTTACAAACAATGTAATAATAG 719

C1GALT1-cDNA 781 ATTTGGAAGAAACCTTTTCATCCCTTTGTGTCAGAAACCATTTAATTAAGGTTATCTACCTAGAACTTTTGGTACTGGAATTACAA 870  
pC1GALT-1 781 ACTGGAAGAAACCTTTTCATCCCTTTGTGTCAGAAACCATTTAATTAAGGTTATCTACCTAGAACCTTTTGGTACTGGAATTACAA 869  
pC1GALT-2 786 ACTGGAAGAAACCTTTTCATCCCTTTATGTCAGAAAC--TTTAATTAAGGTTATCTACCTAGAACCTTTTGGTACTGGAATTACAA 872  
pC1GALT-3 794 ACTGGAAGAAAGACTTTTCATCCCTTTTAGCAAAACGCTAATTAATGAGTTTCTACCTAGAACCTTTTGGTACTGGAATTACAA 881  
pC1GALT-4 720 ACTG-CCTTGTATGTTATCTTC-----ATCTACATGGTTATTTTGGTATGACCTGCTTATTGAGATATCAGTAAACAA 796

C1GALT1-cDNA 871 TAT-TATCCTCCTGTAGAGG-GTCCTGGTTGCTGCTCTCTGATCTGCAAGTTTCTTTTCACTATGTTGATTCTACAACCATGTATGAGTTAG 958  
pC1GALT-1 870 TAT-TATCCTCCTGTAGAGG-GTCCTGGTTGCTGTTCTGATCTGCAAGTTTCTTTTCACTATGTTGATTCTACAACCTGTATGTGTTAG 957  
pC1GALT-2 873 TGT-TATCCTCCTGTAGAGG-GTCCTGGTTCTTGTGTTCTGATCTGCAAGTTTCTTTTCACTAGTTGATTCTACAACCTATGCTGTAGTTAG 960  
pC1GALT-3 882 TAT-TATCCTCTCTGTAGAGG-ATCCAGGTTGCTGTTCTGATTTTGCATTTCTTTTCACTATGTTGATGCTACAGCTATATATGTTAG 969  
pC1GALT-4 797 TAAGAAATGAGATCTCAAGATTAAGAAAGAAATTAACATATCTGAGGAAGGAACTTCT--GCAATTTCTTATATAGAACACCATGTCCCATGAGG 884

III

C1GALT1-cDNA 959 AATACCTCGTTTATCATCTTCGTCATATGGTTATTTATACAGATATCAACCTTACCTGAACTATACTAAAGGAAATTAGTCAAG 1048  
pC1GALT-1 958 AATACCTCGTTTATCATCTTCATCCATATGGTTATTTATACAGATATCAAGCTTACCTGAAATATACTAAAGGAAATTAGTCAAG 1047  
pC1GALT-2 961 GATACCTCATTTATCATCTGCTGCTTATATGATTTATGATACAGATATCAACCTGCTTACCTGAAATATACTAAAGGAAATTAGTCAAG 1050  
pC1GALT-3 970 AATACCTCATTTATCATCTTATCTTATGGTTATTTATACAGATATCAACCTGCTTACCTGAAAGATCTGTTAAAGGAAATAGCAAG 1059  
pC1GALT-4 885 AGTGAATTTGTATGTGAACAT-TTCTATCTTAAATCTTCTATATGACTGGCTG--TACCTGAACTGTGTAATCATGAGTTATGA--AGT 967

C1GALT1-cDNA 1049 CAAACAAAAATGAAGATACAAAAGTGAAGTTAGGAAATCCTTGA---- 1092  
pC1GALT-1 1048 CAAACAAAAATGAAGGACAAAAGTGAAGTCAAGAAACCTTGA---- 1091  
pC1GALT-2 1051 CAAACAAAAATGAAGGACAAAAGTGAAGTTAGGAAACCTTGA---- 1094  
pC1GALT-3 1060 CAAACAAAAATGAAGATCAAAAAGTGAAGCTTAAGAAACCTTCAAAAGA 1107  
pC1GALT-4 968 CTGTTAAAAATACCTTTTATTTTTTGGGACGGAATTTCACTCT----- 1008

**Supplemental Figure 4.** Alignment of human *T-synthase* pseudogenes to the cDNA of the functional gene. The homologous region of the human pseudogenes, *pC1GALT-1*, -2, -3, and -4 were aligned to the open reading frame of the cDNA of the functional T-synthase (*C1GALT1*) using the MacVector9.5 DNA analyzing program. The numbering of nucleotides is shown on both sides of the sequence. The boundaries of the three exons (Exon I, II, and III) are indicated by arrows. Identical sequences are boxed.

## Supporting information: Protein Recognition by an Ensemble of Fluorescent DNA G-Quadruplexes

David Margulies and Andrew D. Hamilton\*

**1. General.**  $^1\text{H}$  NMR spectra were acquired on Bruker Advance DRX 400 series spectrophotometers at 400 MHz. Fluorescence measurements were carried out using Hitachi F-4500 spectrometer with excitation slit 10 nm. Formation of quadruplexes was verified by circular dichroism spectrophotometry using an Aviv62DS spectropolarimeter with a 1 mm pathlength cuvette. In all cases data was subtracted from the spectra of a solution containing only buffer or water. A single drop fluorescence measurement was performed using NanoDrop 3300 fluorospectrometer. Reagents, salts, and proteins were purchased from Sigma-Aldrich, except lipase that was purchased from MP biomedical. Proteins were dissolved in buffer or water, desalted using Bio spin 30 tris columns and their concentrations were determined using their molecular weights and molar absorption coefficients at 276-280 nm. For mouse-MBP and unfractionated whole histone, average values were taken. Oligonucleotides modified with fluorescein, tanra and *N*-hydroxysuccinimide ester were synthesized by the W. M. Keck Foundation Biotechnology Resource Laboratory located at Yale University, using standard automated solid phase synthesis. Concentration of the oligonucleotide was obtained from UV measurements at 260 nm using an Agilent A453 spectropolarimeter, based on their respective electronic absorption at 260 nm and their molar extinction coefficients obtained by nearest neighbor calculations. Extinction coefficients of the dyes were taken as 38,800  $\text{M}^{-1}\text{cm}^{-1}$  fluorescein-dT, 32,300  $\text{M}^{-1}\text{cm}^{-1}$  tanra (Glenresearch), and 22300  $\text{M}^{-1}\text{cm}^{-1}$  for pyrene. Oligonucleotides were characterized by Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) spectrometry (Applied Biosystems Voyager -DE PRO Workstation). Chromatographic separation was achieved by reverse-phase HPLC on a Varian Prostar equipped with a Timberline TL-105 column heater.

**2. Preparation of oligonucleotides.** Synthesis of amino modified pyrene: 1-pyrene carboxylic acid (100 mg, 0.4 mmol) was dissolved in oxalyl chloride (0.7 mL, 8.12 mmol) and DCM (4 mL). Two drops of DMF were added and the solution was stirred under nitrogen overnight. The solvent was evaporated, DIPEA (0.1 mL, 0.6 mmol) and *N*-Boc-1,5-diaminopentane (0.43 mL, 2 mmol) in DCM were added and the solution was stirred for 18 hours. The solvent was removed; the residue was dissolved in DCM, washed several times with 1N  $\text{NaHCO}_3$  and 0.5N HCl, dried and evaporated. The crude product was purified by flash chromatography using 2.5% methanol in DCM as eluents. It was dissolved in DCM (5 mL) followed by addition of TFA (5 mL) and 2 drops of water to the solution. After a few hours the solvent was removed. TFA was co-evaporated with DCM and methanol for several times, followed by evaporation of the solvent under high vacuum to afford the amino modified pyrene in 63% yield.  $^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  = 1.58 (m, 2H,  $\text{CH}_2\text{CH}_2\text{CH}_2$ ), 1.78 (m, 4H,  $\text{CH}_2\text{CH}_2\text{N}$ ,  $\text{NCH}_2\text{CH}_2$ ), 2.98 (t,  $J=7.6$  Hz, 2H,  $\text{CH}_2\text{NH}_2$ ), 3.57 (t,  $J=7.2$  Hz, 2H,  $\text{CH}_2\text{NHCO}$ ), 8.06-8.27 (m, 8H, Pyr), 8.43 (d,  $J=8.0$  Hz, 1H, Pyr). *MS-ES*+: ( $m/z$ ) 331.05.

Synthesis of pyrene appended ODN: to a vial containing 6 mg of amine 1 mL of anhydrous DMF, and 100  $\mu\text{L}$  anhydrous DIPEA were added. The resulting solution was transferred into a syringe and introduced into a cartridge containing a solid support coupled to *N*-hydroxysuccinimide ester (NHS ester) modified oligonucleotide (1.0  $\mu\text{mol}$  scale). The amine solution was pushed through the cartridge several times and then the resin-linked DNA/amine mixture was agitated overnight. The amine solution was then removed and the cartridge was washed three times with 1mL DMF followed by 1mL aliquots of HPLC grade acetonitrile. The resin linked DNA was dried by introducing argon flow through the cartridge for 1 hr. Cleavage from the resin and global deprotection was achieved by introducing a 30% solution of  $\text{NH}_4\text{OH}$  (3 mL) and incubating the solution overnight at 55°C. The solvent was evaporated and the solution was purified by gel filtration using Microspin G-25 Columns (Amersham Biosciences). Further purification of pyrene appended oligonucleotides was achieved using a Varian PRP-1 reverse-phase HPLC column that was maintained at 65°C using a heat jacket. The sample was heated to 95°C for 10 min, and then rapidly introduced into the HPLC injector. The solvent system used was A: 0.1M TEAA, 5% ACN, B: 100% ACN. The pyrene modified ODN was checked by MALDI-TOF. Calculated monoanion; 2864.07, observed: 2862.4. Commercially available fluorescein and tanra appended ODNs were dissolved in water and further purified by gel filtration using Microspin G-25 Columns (Amersham Biosciences). Calculated fluorescein monoanion; 3008.15, observed: 3007.2. Calculated tanra monoanion: 3063.29, observed: 3062.3.

**3. Formation of quadruplex ensembles and protein detection.** Stock solutions of PFT quadruplex ensemble were prepared by mixing the three ODNs in a minimum amount of water (<10 $\mu\text{L}$ ), followed by addition of 130  $\mu\text{L}$  of 10 mM Tris-HCl, 80 mM KCl, pH 7.3, such that the final concentrations of ODNs resulted in 40  $\mu\text{M}$  P and 44  $\mu\text{M}$  F and 44  $\mu\text{M}$  T (32  $\mu\text{M}$  quadruplexes). After fastening the cap and sealing with parafilm quadruplex formation was initiated by heating the solution at 97°C for 15min, followed by slow cooling to room temperature and 48h incubation at 4°C. The PFT\* ensemble was prepared by heating a water solution, containing the same ODNs concentrations, to 97°C for 15, followed by slow cooling to room temperature. The solution was then heated again for 15 minutes, removed from the heater and rapidly evaporated to dryness under vacuum. The solid was dissolved in 10mM Tris-HCl, 80mM KCl, pH 7.3, and quadruplex formation was achieved by following by the same protocol as for PFT. Before each fluorescence titration, fresh quadruplex stock solutions were diluted by 70 folds with buffer. In a single microliter measurement, a drop of buffer is first measured as blank, after which drops of PFT or PFT loaded with a protein are measured.

**4. Principal component analysis.** Analysis of the emission spectra was performed by XLSTAT 7.5.3 - Principal Component Analysis (PCA) on 57 data points corresponding to the change in emission  $[(I_{\text{protein}} - I_{\text{ref}}) / I_{\text{ref}}] * 100$  at 370-650 nm every 5 nm.